**PROPOSTAS PARA DESENVOLVIMENTO DE FUNÇÃO NA LINGUAGEM R**

**1ª PROPOSTA**

**Nome da função: imc.testet**

Entrada dos dados por vetores x para altura(m) e y para peso (kg) ou Dataframe contendo colunas com altura(m) e peso (kg).

1) Calcular o índice de massa corporal (IMC) a partir dos dados dos pacientes (altura e peso) utilizados para este cálculo.

2) Fazer um histograma e boxplot para verificar como estão distribuídos os dados gerais.

Pressupostos para realizar **Teste t:** os dados devem seguir uma distribuição normal ou ter uma amostra grande acima de 30.

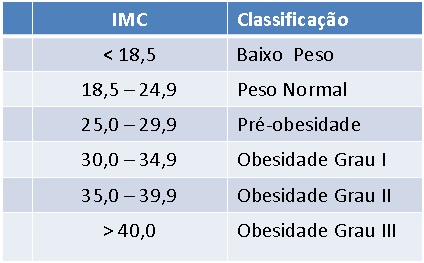
Fazer um bloxplot com as faixas de idade : 20 a 30, 30 a 40, 40 a 50, 50 a 60 e mais de 60.

Fazer um histograma e boxplot para verificar como estão distribuídos os dados por faixa etária e sexo.

3) **Teste t** (Unicaudal) para a média aritmética dos dados da amostra, assim podemos classificar em peso normal e peso fora do normal.

μ é o valor do imc da população que encontramos na literatura.

Fazer um **Teste t** para os dados de imc, sendo Ho μ ≤ 24,9 kg/m² e H1 μ ˃ 24,9 kg/m²



Fonte: <https://www.hospitalsiriolibanes.org.br/hospital/especialidades/nucleo-obesidade-transtornos-alimentares/Paginas/calcule-seu-imc.aspx>

4) Visualizando os dados de imc separados.

Se os valores do imc forem ≤ 24,9 kg/m² realizar os gráficos boxplot e histograma para esta faixa de imc e exibir junto com um sumário dos dados.

Se os valores do imc NÃO forem ≤ 24,9 kg/m² e idade ˃ 40 anos realizar os gráficos boxplot e histograma para esta faixa de imc e idade, exibir junto com um sumário dos dados.

**Observações:** Na função foi colocado um ponte de corte para separar pessoas com mais de 40 anos.

**Exemplo:** O câncer de mama é o câncer de maior incidência e prevalência nas mulheres. Há estudos que mostram que depois da menopausa o sobrepeso ou obesidade é um fator importante no prognóstico de tratamento. Assim, fazer uma análise exploratória com faixas de 10 anos seria importante para ter uma visão global dos dados.

**2ª PROPOSTA**

**Nome da função: imc.testes**

Entrada dos dados por vetores x para altura(m) e y para peso (kg) ou Dataframe contendo colunas com altura(m) e peso (kg).

1) Calcular o índice de massa corporal (IMC) a partir dos dados dos pacientes (altura e peso) utilizados para este cálculo.

2) Fazer um histograma e boxplot para verificar como estão distribuídos os dados gerais.

Pressupostos para realizar **Teste t:** os dados devem seguir uma distribuição normal ou ter uma amostra grande acima de 30.

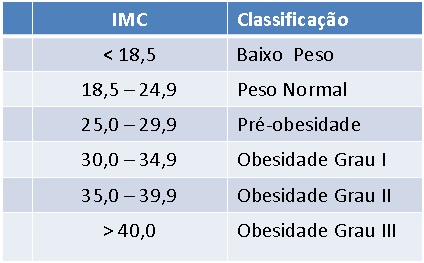
Fazer um bloxplot com as faixas de idade : 20 a 30, 30 a 40, 40 a 50, 50 a 60 e mais de 60.

3) Fazer um histograma e boxplot para verificar como estão distribuídos os dados por faixa etária e sexo.

4) **Teste t** (Unicaudal) para a média aritmética dos dados da amostra, assim podemos classificar em peso normal e peso fora do normal.

μ é o valor do imc da população que encontramos na literatura.

Fazer um **Teste t** para os dados de imc, sendo Ho μ ≤ 24,9 kg/m² e H1 μ ˃ 24,9 kg/m²



Fonte: <https://www.hospitalsiriolibanes.org.br/hospital/especialidades/nucleo-obesidade-transtornos-alimentares/Paginas/calcule-seu-imc.aspx>

5) Teste F para variância de duas amostras

Se o valor do imc for μ ˃ 24,9 kg/m² realizar um Teste F para saber a variância nos dados da amostra com valor superior ao imc de peso normal em relação a amostra com peso normal.

Se o imc não for μ ˃ 24,9 kg/m², ou seja, for μ ≤ 24,9 kg/m², realizar somente os gráficos boxplot e histograma para esta faixa de imc e exibir junto com o valor do imc e um sumário dos dados.

**Observações:** Na função foi colocado um ponte de corte para separar pessoas com mais de 40 anos.

**Exemplo:** O câncer de mama é o câncer de maior incidência e prevalência nas mulheres. Há estudos que mostram que depois da menopausa o sobrepeso ou obesidade é um fator importante no prognóstico de tratamento. Assim, fazer uma análise exploratória com faixas de 10 anos seria importante para ter uma visão global dos dados.

**3ª PROPOSTA**

**Nome da função: imc.anova**

Entrada dos dados por vetores x para altura(m) e y para peso (kg) ou Dataframe contendo colunas com altura(m) e peso (kg).

1) Calcular o índice de massa corporal (IMC) a partir dos dados dos pacientes (altura e peso) utilizados para este cálculo e realizar uma análise exploratória dos dados.

2) Fazer um histograma e boxplot para verificar como estão distribuídos os dados gerais.

Pressupostos para realizar **ANOVA:** os dados devem seguir uma distribuição normal ou ter uma amostra grande acima de 30.

Fazer um bloxplot com as faixas de idade: 20 a 30, 30 a 40, 40 a 50, 50 a 60 e mais de 60.

3) Fazer um histograma e boxplot para verificar como estão distribuídos os dados por faixa etária e sexo.

4) ANOVA

c= número de grupos = numero de faixas de idade= 5

n= tamanho da amostra

Selecionar um tamanho de amostra igual para cada grupo (faixa etárias), uma vez que é um pressuposto para ter um teste anova adequado.

Realizar teste ANOVA

**Se F estat ˃ Fα (α=5%)**

Exibir summary do teste Anova e mensagem: Existe diferença entre as médias aritméticas dos grupos.

**Se F estat ˂ Fα (α=5%)**

Exibir summary do teste Anova e mensagem: Não há diferenças entre as médias aritméticas dos grupos

**4ª PROPOSTA**

**Nome da função: prevencao**

Calcular o índice de massa corporal (imc) a partir dos dados de altura(m)e peso(kg) dos pacientes e realizar uma análise exploratória dos dados dos pacientes em relação a gordura corporal, atividade física, alimentos de origem vegetal, alimentos de origem animal (limite de consumo de carne vermelha) e bebidas alcoólicas (limite de consumo de bebida alcóolica).

Separar os dados de homens e mulheres do dataframe de entrada.

Realizar um gráfico boxplot com faixas de 10 anos para ter uma visão mais específica dos dados por sexo.

As faixas seriam [20,30[ , [30,40[ , [40,50[ , [50, 60[ e 60 ou mais.

Calcular a média e mediana para cada uma das colunas: índice de massa corporal (imc), atividade física ( valor em minutos por dia), alimentos de origem vegetal (quantos vezes por semana), alimentos de origem animal (quantas vezes por semana), e bebidas (quantas vezes por semana)**.**

Fazer bloxplot de cada variável da coluna por faixa etária.

Fazer bloxplot de cada variável da coluna por sexo.

Exibir os gráficos

**Exemplo:** Estes parâmetros seguem as recomendações gerais do relatório sobre alimentação e câncer do WCRF/AICR de 2007.

**5ª PROPOSTA**

**Regressão Linear**

A função terá um limite de 6 variáveis diferentes.

1) Plotar a variável resposta (dependente) com cada variável preditora (independente).

2) Fazer uma regressão linear simples de cada variável resposta (dependente) com cada variável preditora (independente)

3) Plotar o gráfico de cada regressão linear simples

4) Mostrar o resultado de cada coeficiente de cada regressão linear simples

5) Fazer um teste ANOVA e comparar cada variável com o modelo nulo

6) Summary de cada regressão linear simples

7) Organizar os Multiple R-squared em ordem decrescente, selecionar as 3 maiores.

As variáveis que tiverem maior Multiple R-squared serão incorporadas ao modelo de regressão múltipla com no máximo 3 variáveis preditoras e realizar os modelos com interação. Não será feitos mudança na ordem das variáveis preditoras, serão colocadas conforme a ordem decrescente.